PCT/GB 87 00039

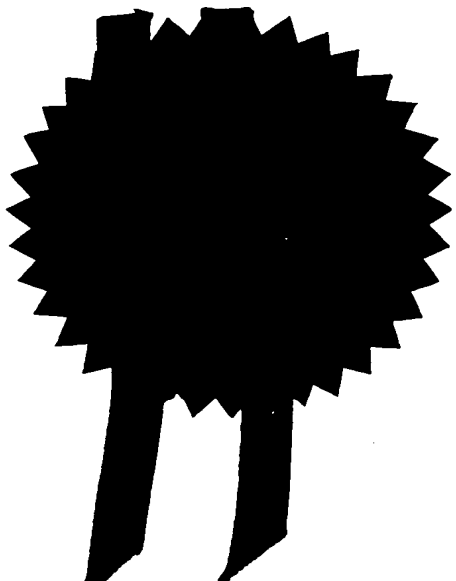
18 Rec'd PCT/PTO 29 JUL 1987

THE PATENT OFFICE
STATE HOUSE
66–71 HIGH HOLBORN
LONDON WC1R 4TP

I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents, has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.



Please turn over

23 JAN 1986

PATENTS ACT 1977

PATENTS FORM No. 1/77 (Revised 1982)
(Rules 16, 19)

The Comptroller
The Patent Office
25 Southampton Buildings
London, WC2A 1AY

1986
01597

24/01/86 C8627 PAT*** 10.00

REQUEST FOR GRANT OF A PATENT

8601597

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

I Applicant's or Agent's Reference *(Please insert if available)* PA 108

II Title of Invention NUCLEOTIDE SEQUENCES AND USES THEREOF

III Applicant or Applicants *(See note 2)*

Name (First or only applicant) RICHARD HARRIS WILSON

Country United Kingdom State ADP Code No.

Address 48 GALBRAITH DRIVE, MILNGAVIE, GLASGOW, G 62 6NE
SCOTLAND

Name (of second applicant, if more than one)

Country State

Address

IV Inventor (see note 3)

~~(a) The applicant(s) is/are the sole/joint inventor(s)~~
~~or~~
(b) A statement on Patents Form No 7/77 ~~is~~/will be furnished

V Name of Agent (if any) *(See note 4)* CARPAELS & RAMSFORD ~~Mr. P. E. Crawley~~ ADP CODE NO

VI Address for Service (See note 5) ~~Celltech Limited, 244-250 Bath Road, Slough, Berkshire, SL1 4DY Tel: Slough (0753) 36162~~ 43 BLOOMSBURY SQUARE LONDON WC1A 2RA

~~VII Declaration of Priority (See note 6)~~

Country Filing date File number

VIII The Application claims an earlier date under Section 8(3), 12(6), 15(4), or 37(4) (See note 7)

~~Earlier application or patent number and filing date~~

PA 108

Nucleotide Sequences and Uses Thereof

The present invention relates to recombinant DNA technology. In particular the invention relates to nucleotide sequences which code for glutamine synthetase (L-glutamate: ammonia ligase [ADP-forming], (EC 6.3.1.2))(GS) and to uses of such nucleotide sequences.

Glutamine synthetase (GS) is a universal housekeeping enzyme responsible for the synthesis of glutamine from glutamate and ammonia using the hydrolysis of ATP to ADP and phosphate to drive the reaction. As such it is involved in the integration of nitrogen metabolism with energy metabolism via the TCA cycle, glutamine being the major respiratory fuel for a wide variety, possibly the majority, of cell types.

GS is found at low levels (0.1% - 0.01% of soluble protein) in most higher vertebrate cells and is found at higher levels (>1% of total protein) in certain specialised cell types such as hepatocytes, adipocytes and glial cells. A variety of regulatory signals affect GS levels within cells, including glucocorticoid steroids and cAMP, and medium glutamine appears to regulate GS levels post translationally via ADP ribosylation.

GS from all sources is subject to inhibition by methionine sulphoximine (Msx), this compound appearing to act as a transition state analogue of the catalytic process. Extensively amplified GS genes have been obtained (Wilson R. H., Heredity 49, 181 (1982) and Young A. P. & Ringold G. M., (1983) J. Biol. Chem., 258, 11260-11266) in variants of certain mammalian cell lines selected for Msx resistance. Recently Sanders and Wilson (Sanders P. G. & Wilson R. H. The EMBO Journal, Vol 3, No. 1, pp65-71, 1984) have described the cloning of an 8.2 kb Bgl II fragment containing DNA coding for GS from the genome of an Msx resistant Chinese hamster overy (CHO) cell line KGIMS. However this fragment does not appear to contain a complete GS gene and it was not sequenced.

We have now further applied the techniques of recombinant DNA technology to GS and have obtained recombinant cDNA clones corresponding to the whole of the mRNA coding for a GS from gene amplified CHO cells. These clones have been sequenced enabling us to determine the cDNA sequence corresponding to the coding portion of the mRNA of this GS. This has further enabled us, not only to predict the full length amino acid sequence of this GS, but also to make effective use of this DNA sequence in a number of recombinant DNA applications. This work provides, for the first time a full length gene sequence coding for an eucaryotic GS.

Accordingly in a first aspect the invention provides a nucleotide sequence coding for the complete amino acid sequence of an eucaryotic GS.

The nucleotide sequence of the first aspect of the invention typically comprises a DNA sequence coding for the complete amino sequence of a mammalian GS. In particular the DNA sequence comprises a DNA sequence, preferably a cDNA sequence, coding for the GS of a rodent such as a mouse, rat or especially a hamster. For instance, the sequence may be a hamster GS coding sequence substantially as hereinafter shown in Figure 2. It will be appreciated, however that the DNA sequence of Figure 2 (or the corresponding mRNA sequence) or parts thereof may be used as hybridisation probes for obtaining GS coding sequences from other sources including cells of different animal species, using DNA hybridisation techniques. Thus the invention includes DNA sequences which hybridise under appropriate conditions, e.g. conditions of high stringency, with the sequence of Figure 2 or part thereof or corresponding RNA sequences.

Furthermore nucleotide sequences coding for GS e.g. the sequence of Figure 2, parts thereof, or RNA corresponding thereto may be used in diagnostic or medical applications; for instance, in those diseases states, such as those associated with certain tumours, where GS levels are altered.

Generally however the DNA sequence of the invention may be used in a variety of recombinant DNA applications.

Thus in a second aspect the invention provides a vector containing a DNA sequence according to the first aspect of the invention.

Furthermore, in a third aspect the invention provides host cells transformed with a vector containing a DNA sequence according to the first aspect of the invention.

A GS DNA sequence according to the invention may be used in the co-amplification of non-selected genes. In such applications the GS gene and the non-selected gene may be present either in the same vector or may be co-transfected into appropriate host cells in separate vectors.

Also the GS DNA sequence may be used as a dominant selectable marker for introducing genes, e.g. heterologous genes, into appropriate host cells.

Preferably GS may be used for co-amplification of non-selected genes or as a dominant selectable marker in higher eucaryotic host cells, such as mammalian cells e.g. wild type CHO cells.

GS has advantages for use in these applications as compared with other genes; for instance, the DHFR: methotrexate co-amplification system which has been described previously by other workers e.g. Axel et. al. (US Patent No. 4 399 216), in view of particular aspects of glutamine metabolism and the means which are available for regulating GS synthesis and turnover.

Thus various cell lines differ in their endogenous GS activities and in their ability to utilise glutamate in the medium as a substrate for the enzyme.

GS deficient cell lines may be obtained by screening or selection using known techniques, and such cell lines may provide host cells for transfection with GS genes. The GS gene may be used as a dominant selectable marker for transfection of such host cells. In these cells the GS gene may act as a dominant selectable marker for transfection, permitting growth in medium lacking glutamine but containing glutamate.

Furthermore, other cell lines are defective in their ability to take up glutamate, and variants may be selected which are able to grow on medium lacking glutamine but containing glutamate. Such variant cell lines may also be used as host cells for transfection with a GS gene as a dominant selectable marker, even when endogenous GS genes are active.

GS deficient cell lines and variant cell lines as described above, and advantageously also cell lines which are adapted to glutamate utilisation, may be transfected using a GS gene as a dominant selectable marker. Thus the GS gene may endow the transfected host cells with resistance to levels of glutamine analogues, such as Msx, which would be toxic for corresponding non-transfected host cells. Advantageously, the levels of glutamine anologues used for selection are sufficiently high to prevent selection of resistant non-tranfected variants arising from amplification of endogenous (resident) GS genes. In preferred embodiments endogenous GS activity may be reduced or even abolished, for instance, by treatment with dibutyryl-cAMP + theophylline e.g. in the case of 3T3-L1 cells.

Co-amplification of the GS and non-selected genes in transfected cells may be achieved by selection for resistance to progressively increased levels of an inhibitor of GS e.g. phosphinothricin or Msx.

Such GS inhibitors are typically highly specific to GS, cheap to produce and have high solubility in aqueous solution, advantageously permitting high concentrations of the toxic inhibitor to be used.

Further advantages of GS stems from the potentiation of the inhibitory effect of Msx which may be achieved. Thus, for example, Msx may be potentiated by addition of methionine to the culture medium, or alternatively by reducing the glutamate concentration in the medium. In this way relatively low levels of toxic inhibitor may be required to select cells containing very highly amplified genes.

Also, in particularly preferred embodiments, the requirement for Msx may be further reduced in highly amplified cells by placing the GS gene under

the control of a regulatable promoter e.g. a heat shock promoter or a metallothionein promoter. In such embodiments, GS gene expression may be switched on to provide a dominant selectable marker and increase copy number of the GS and non selected genes by amplification, and then down regulated to reduce the requirement for Msx.

Thus in preferred embodiments of the second and third aspects of the invention the GS DNA sequence is under the control of a regulatable promoter.

Further advantages for use of GS arise from the possibility of using culture medium which is substantially free of (or at least significantly reduced in respect of) the GS inhibitor once maximal and stable amplification has been achieved. Thus in preferred embodiments the large amount of metabolically active GS which the transfected, amplified cell is producing may be reduced by addition of glutamine or glutamine analogues to the culture medium. This is possible, becuause glutamine down regultes GS activity in many cultured cell lines, probably as a result of increased enzyme turnover. Under conditions of low or zero inhibitor concentration transformed cell line stability is increased since selection pressure for non-productive revertants is non-existant or markedly reduced.

The various advantages of GS for co-amplification, as outlined above, may provide a more desirable co-amplification system than co-amplification system, such as the DHFR-methotrexate system, which have been proposed previously. Thus, for example, DHFR is typically constitutively expressed by cells at relatively constant levels and it is usually necessary to use DHFR$^-$ host cells, and such cells may not be good producers of heterologous gene products. In comparison endogeneous (resident) GS gene expression may be controlled by a variety of means permitting advantageous selection of incoming over resident GS genes. Furthermore requirements for GS inhibitor by transfected, amplified cell lines may be adjusted as described above. Generally GS may provide an advantageously flexible and

controllable coamplification system.

Furthermore the GS coding sequences of the invention may be used to cure cells from glutamine dependance; for instance, to overcome growth limitation due to ammonia production.

For example, a GS gene under the control of a suitable promoter may be used to transfect cell lines, and such transfected cell lines advantageously express high levels of GS activity and are subject to lower levels of ammonia accumulation when grown in medium lacking glutamine, and thus may grow to higher cell densities than corresponding untransfected cell lines.

The invention is further described by way of illustration in the following examples which relate to the determination of the nucleotide sequence coding for GS from gene amplified CHO cells and which refer to the accompanying diagrams in which

Figure 1 shows restriction maps of cloned cDNAs coding for the GS;

Figure 2 shows the cDNA sequence and corresponding amino acid sequence of a GS from gene amplified CHO cells;

Figure 3 shows the mRNA sequence of the GS in the region of a polyadenylation site and, for the sake of comparison, the corresponding region of human U4 RNA, and

Figure 4 shows the GS cDNA sequence of Figure 2.

In the nucleotide sequences of Figure 2, 3 and 4 and elsewhere in the present description:

U - denotes a uridine residue

G - denotes a guanosine nucleotide reside

T - denotes a thymidine nucleotide residue

A - denotes a adenosine nucleotide residue, and

C - denotes a cytosine nucleotide residue.

*** denotes a termination codon.

- denotes an unknown nucleotide residue.

In the amino acid sequences of Figure 2 and elsewhere in the present description:

A - denotes an alanine residue

C - denotes a cysteine residue

D - denotes an aspartic acid residue

E - denotes an glutamic acid residue

F - denotes a phenylalanine residue

G - denotes a glycine residue

H - denotes a histidine residue

I - denotes an isoleucine residue

K - denotes a lysine residue

L - denotes a leucine residue

M - denotes a methionine residue

N - denotes an asparagine residue

P - denotes a proline residue

Q - denotes a glutamine residue

R - denotes an arginine residue

S - denotes a serine residue

T - denotes a threonine residue

V - denotes a valine residue

W - denotes a tryptophan residue

Y - denotes a tyrosine residue, and

X - denotes an unknown amino acid residue

EXAMPLE

MATERIALS AND METHODS

Growth of cells and isolation of mRNA

Growth of cells and isolation of mRNA was done as previously (13).

Isolation of Genomic GS Subclones

pGS113 is a 3.5kb HindIII fragment containing the 3' end of the GS gene subcloned from pGS1 into pUC9 (13). pGS2335 is a BamHI-EcoRI subclone of a λL47 recombinant (14) selected from a clone bank prepared by cloning a Sau3A partial digest of GS-amplified CHO cell DNA into the BamHI site of λL47 and selecting for hybridisation to pGS1 (R.H. Wilson, P.G. Sanders and B.E. Hayward, in preparation).

cDNA Cloning and Sequencing

cDNA libraries were made in pBR322 and λgt10 using standard procedures. Messenger RNA was converted to cDNA using oligodT primed reverse transcriptase, and ds DNA made by the RNase H procedure (15). The ds DNA was either tailed with C residues (16), annealed to G-tailed pBR322 (obtained from BRL) and transformed into E.coli DH1, or methylated and ligated to EcoRI linkers. Linkered DNA was digested with EcoRI and linkers removed by Sephadex G75 chromatography in TNES (0.14 M NaCl, 0.01 M Tris pH 7.6 0.000 M EDTA 0.1% SDS). Linkered DNA in the excluded volume was recovered by ethanol precipitation and annealed to EcoRI cut λgt10 DNA. Following in vitro packaging, recombinant phage was plated on a high frequency lysogeny strain of E.coli (Hf1) (17).

About 5000 colonies and 20000 plaques were screened on nitrocellulose filters using nick-translated probes derived from pUC subclones of GS genomic sequences. A 1kb EcoRI-Bg1II fragment from pGS2335 was used as a 5' probe, and the entire 3.5kb HindIII fragment of pGS113 was used as a 3' probe. Plasmids from positive colonies were analysed by restriction digestion of small-scale preparations of DNA and the longest clone (pGSC45) selected for further analysis.

Positive λ clones were plaque purified, grown up in 500 ml of E.coli C600 liquid culture, and the phage purified on CsCl step gradients. λ DNA was prepared by formamide extraction (18). Clones with the longest inserts were indentified by EcoRI digestion and inserts subcloned into pAT153 and M13mp phage for further analysis and sequencing (19).

RESULTS AND DISCUSSION

cDNA Cloning

The availability of mRNA from a relatively abundant source (Msx amplified CHO cells) and plasmid sub-clones of λ phage GS gene recombinants for use as probes contributed to the success of the cloning strategy.

Two cDNA libraries were made; C-tailed cDNA was annealed to G tailed pBR322 and transformed into E.coli DH1, and EcoRI linkered and methylated cDNA was annealed to EcoRI cut λgt10 DNA, and after in vitro packaging recombinant phage were plated on Hf1 E.coli.

The colonies or plaques were screened first with a probe derived from the 5' end of the GS gene. Positive colonies or plaques from this analysis were picked and rescreened with a longer probe covering most of the 3' end of the GS gene. In this way it was anticipated that clones with long or possibly full length inserts would be selected and that tedious rescreening for 5' ends would be avoided. Several plasmid clones and two λgt10 recombinants were derived by this means. Further analysis of one of the plasmid clones (pGSC45) by restriction enzyme digestion and partial sequencing revealed that it had an insert of about 2.8kb and a polyA sequence at the 3' end. Northern blots indicate that a major mRNA for GS is about this size (13), so the insert in pGSC45 was potentially a full length copy of this mRNA. The two λ clones (λgs 1.1 and λgs 5.21) had inserts of 1450 bp and 1170 bp respectively. Restriction maps and alignment of the cDNA inserts in pGSC45, λgs 1.1 and λgs 5.21 are shown in Fig. 1. It is clear that the inserts in the λ clones are considerably

shorter at the 3' end than the plasmid clone and may represent cDNA copies of one of the minor mRNAs. The insert in λgs 1.1 extends some 200 base pairs at the 5' end.

Analysis of the Nucleotide Sequence

The nucleotide sequence of the mRNA coding for glutamine synthetase (Fig. 2) was obtained from M13 subclones of pGSC45 and EcoRI subclones of λgs 1.1 and λgs 5.21 (Fig. 1). Some confirmatory sequence was also obtained from the genomic clone pGS1 (13). Primer extension off GS mRNA with an oligonucleotide complementary to nucleotides 147-166 gave a major extension product of 166 nucleotides. This shows that pGSC45 only lacks six or seven nucleotides from the 5' end of the mRNA (Fig. 2). Nucleotide sequencing of the primer extended product by Maxam-Gilbert sequencing (20) confirmed this although the first two bases could not be determined.

The 5' noncoding sequence of the GS cDNA is 146 bases with a G+C proportion of 64.6% (compared with the coding portion at 53.1% G+C). The 5' noncoding region of the mRNA can assume a conformation with a free-energy of -53.3 Kcal as calculated by the program of Zuker and Stiegler (21). This structure folds into two extended stem-loop structures centred on bases 43 and 107, but leaves the 10 bases upstream of the initiation codon free (Fig. 2). Two regions of the sequence not involved in the proposed structure have some homology to the 3' end of 18S rRNA, namely bases 96-102 and 138-143 (22). Despite being nearly 1kb shorter than the insert in pGSC45, the 3' end of λgs 5.21 also contained a polyA sequence, suggesting that this clone was derived from the minor mRNA of c 1.4kb found in amplified CHO cells (13). A 3' untranslated region of 108 nucleotides lies upstream of this site but there is no AATAAA consensus contained there. However, the region does show several stretches of complementarity to the 5' end of U4 RNA (23), the best being that which flanks the polyadenylation site and presents it is a structure similar to those proposed for polyadenylation by Berget (24).

This is shown in Fig. 3.

Sequences at the 5' end of λgs 1.1, which is some 200 bases longer at the 5' end than pGSC45, showed considerable inverted homology to sequences at the 3' end of this clone (which was about 150 bases shorter at the 3' end than λgs 5.21, see Fig. 1). These additional sequences are probably cloning artefacts, arising during second strand synthesis (27) due to nucleotides 6-1 priming DNA synthesis via their complementarity to nucleotides 1132-1137 despite the fact that the RNase H procedure (15) was used. It cannot be excluded that the duplication arises from transcription of a modified GS gene, producing a modified mRNA which has been subsequently cloned, although the primer extension results did not suggest that there was any major mRNA species with a 5' end longer than 166 nucleotides.

The precise relationship between the multiple mRNA species found in CHO cells with amplified GS genes, those found in wild type CHO cells and the cDNA clones, will require further analysis. In similar cells with amplified dihydrofolate reductase genes mRNA species with variant transcription initiation and polyadenylation sites have been observed (25, 26). Preliminary comparisons of the organisation of the DNA in clones covering the complete GS gene and that indicated by Southern blots of wild type CHO cell DNA, suggest that no detectable rearrangements of the coding sequence have occurred (M. Macdonald and R.W. unpublished results).

Predicted Amino Acid Sequence

The predicted amino acid sequence for CHO glutamine synthetase in shown in Fig 2. The NH_2 terminus was identified by homology with the NH_2 terminal peptide found in bovine brain glutamine synthetase (28). The initiating AUG follows a precise CCACC upstream consensus sequence found for true initiation codons (29) and is followed by a purine (i.e. CCACCATGG). Another AUG codon at position 14 is not in a favourable context by the same criteria and is followed by a termination codon in

frame 21 nucleotides downstream. The predicted amino aicd composition of the GS protein gives a molecular weight of 41,964 (not allowing for N-terminal acetylation (28) or other post-translational modifications), in agreement with other estimates (13,30). The basic nature of the protein is reflected in the excess of arginine [23], histidine [13] and lysine [20] residues over those of aspartate [18] and glutamate [27].

The codon usage of the Chinese hamster GS mRNA conforms well to a mammalian consensus (31). Exceptions are that the CCY proline codons are favoured (22/1 observed, 15/8 expected), and the CGN arginine codons are unduly favoured over the AGR codons (21/3 observed, 13/11 expected). The biased codons are not clustered in the mRNA.

Our amino acid sequence shows excellent homology with bovine and other GS sequences obtained by peptide sequencing indicative of an accurate DNA sequence (28, 32). The amino acid sequence allows the ordering of all the cyanogen bromide peptides and most of the tryptic peptides published for bovine GS (28). Many of the differences between the bovine and hamster sequences can be identified as being due to single base changes usually leading to the conservative substitution of an amino acid. However, at two regions the amino acid sequences differ significantly; bovine GS residues 52 - 60 would seem to be better placed as 106 - 114, and bovine residues 277 - 281 would be better placed as 305 - 309, substituting the bovine tryptic peptide T IX - B-1 as 276 - 278. Additionally, we would reverse lysine 105 and arginine 106 of the bovine sequence in order to improve homology and to locate bovine peptide T X-B (28).

The CHO GS sequence shows some homology with the GS sequence from the cyanobacterium Anabaena (33) notably at residues 317-325 (N-R-S-A-S-I-R-P) which are an exact match to Anabaena residues 342-350. In addition, related sequences can be found in glutamine synthetases isolated from plants (34, 35). A more detailed analysis of GS sequence comparisons will be published elsewhere.

The reaction catalysed by GS is complex (1). ATP hydrolysis leads to enzyme-bound gamma-glutamyl-phosphate, which is subject to attack by ammonia to yield glutamine, phosphate and ADP. Consequently, we could expect a tertiary domain structure of GS that might reflect the three different substrate binding sites. However, no obvious homologies were found to the G.X.G.X.X.G. nucleotide-binding domain of many enzymes (36) or to other sequence motifs of ATP-binding enzymes (37). Owing to this, and because of the lack of consensus sequences for arginine-ADP-ribosylation, we are unable to speculate as to the regulatory arginine residue involved in ADP-ribosylation (9). Carbamoyl-phosphate synthetase (CPS) has a similar reaction mechanism to GS (1,38), but comparisons of our GS sequence with that of E.coli CPS (39) show no simple homologies.

Access to complete cDNA clones and genomic clones for Chinese hamster GS has not only allowed the amino acid sequence of glutamine synthetase to be predicted, but will allow a detailed analysis of the position of the introns within the gene and their relationship to the exons coding for the structural domains of the protein.

An active GS transcription unit was constructed from the cDNA clones described above, under the control of promoter and polyadenylation sequences from SV40. This transcription unit was used as a dominant selectable marker to introduce a gene coding for tissue plasminogen activator (tPA) into CHO cells. Co-amplification was achieved by selection in Msx and high levels of tPA production were obtained.

Fig. 1

Restriction maps of the GS specific cDNA inserts in pGSC45, λgs 1.1 and λgs 5.21 clones. As can be seen from the arrows, the nucleotide sequence of the coding region of GS was predominantly obtained from M13 subclones of λgs 1.1 and various regions confirmed using subclones of λgs 5.21 and pGSC45.

Fig. 2

Nucleotide (a:) and predicted amino acid (b:) sequences for the Chinese hamster GS gene, together with the published (28) peptide sequences (c:) and peptide designations (d:) of the bovine brain GS. The sequence (e:) represents the polyadenylation site used in λgs 1.1. Amino acid residues are indicated as their single letter codes; non-homologous bovine residues are indicated in lower case letters. The '^' below base 7 represents the start of the pGSC45 insert and the '----' marker represents the priming sequence in λgs 1.1 complementary to residues 1135-1132. The '>' and '<' symbols represent bases involved in stems of the calculated structure for the 5' untranslated region.

Fig. 3

Homologies between the weak polyadenylation site of the short Chinese hamster mRNA and human U4 RNA. Polyadenylation occurs either before or after the A residue marked with an asterisk.

Fig. 4

Chinese hamster glutamine synthetase (L-glutamate:ammonia ligase (ADP-forming), EC 6.3.1.2) cDNA sequence.

From gene-amplified variant CHO cells. No evidence for any rearrangements in the cDNA sequence compared with wild-type.

1-2	Bases unknown
147-149	Initiation codon
150-1265	Coding sequence (372 a.a.)
1266-1268	Termination codon (UAA)
1377	Inefficient polyadenylation site
1421	End of sequenced region of cDNA
(c.2800)	Polyadenylation site of full-length mRNA

REFERENCES

1. Meister, A. (1980). In Mora, J. and Palacios, R. (Eds). Glutamine metabolism, enzymology and regulation. Academic Press, N.Y., pp. 1-40.
2. Krebs, H. (1980) in Mora J. and Palacios, R. (Eds). Glutamine metabolism, enzymology and regulation. Academic Press, N.Y., pp.319-329.
3. Tiemeier, D.C. and Milman, G. (1972). J. Biol. Chem. 247, 2272-2277.
4. Gebhardt, R. and Mecke, D. (1983). EMBO J. 2, 567-570.
5. Miller, R.E., Hackenburg, R. and Gershman, H. (1978). Proc. Natl. Acad. Sci. USA 75, 1418-1422.
6. Linser, P. and Moscona, A.A. (1979). Proc. Natl. Acad. Sci. USA 76, 6476-6480.
7. Milman, G., Portnoff, L.S. and Tiemeier, D.C. (1975). J. Biol. Chem. 250, 1393-1399.
8. Arad, G., Freikopf, A. and Kulka, R.G. (1976). Cell 8, 59-101.
9. Moss, J., Watkins, P.A., Stanley, S.J. Purnell, M.R. and Kidwell, W.R. (1984). J. Biol. Chem 259, 5100-5104.
10. Ronzio, R.A., Rowe, W.B. and Meister, A. (1969). Biochemistry 8, 1066-1075.
11. Young, A.P. and Ringold, G.M. (1983). J. Biol. Chem. 258, 11260-11266.
12. Wilson, R.H. (1982). Heredity 49, 131.
13. Sanders, P.G. and Wilson, R.H. (1984). EMBO J. 3, 65-71.
14. Loenen, W.A.M. and Brammar, W.J. (1980). Gene 20, 249-259.
15. Gubler, U. and Hoffmann, V. (1983). Gene 25, 263-269.
16. Michelson, A.M. and Orkin, S.H. (1982). J. Biol. Chem. 257, 14773-14782.
17. Huyhn, T.V., Young, R.A. and Davis, R.W. (1985). In DNA cloning techniques II: A practical approach (Ed Glover D.M.), I.R.L. Press, Oxford.
18. Davis, R.W., Botstein, D. and Roth, J.R. (1980). Advanced Bacterial Genetics, Cold Spring Harbor.
19. Maniatis, T., Fritsch, E. and Sambrook, J. (1982). Molecular cloning - A laboratory manual, Cold Spring Harbor.
20. Maxam, A. and Gilbert, W. (1980). Methods in Enzymol. 65, 499-560.
21. Zuker, M. and Stiegler, P. (1981). Nucleic Acids Res. 9, 133-148.
22. Chan, Y.L., Gutell, R., Noller, H.F. and Wool, I.G. (1984). J. Biol Chem. 259, 224-230.
23. Busch, H., Reddy, V.R., Rothblum, L. and Choi, Y.C. (1980). Ann Rev. Biochem. 51, 617-654.
24. Berget, S.M. (1984). Nature 309, 179-182.
25. Masters, J.N. and Attardi, G. (1985). Mol. Cell. Biol. 5, 493-500.
26. Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G., and Kellems, R.E. (1985). Mol. Cell. Biol. 4, 2921-2924.
27. Volckaert, G., Tavernier, J., Derynck, R., Devos, R. and Fiers, W. (1981). Gene 15, 215-233.
28. Johnson, R.J. and Piskiewicz, D. (1985). Biochim Biophys Acta 827, 439-446.
29. Kozak, M. (1984). Nucleic Acids Res. 12, 857-872.
30. Tate, S.S. and Meister, A. (1973). In. Prusiner, S. and Stadtman, E.R. (Eds). The enzymes of glutamine metabolism. Academic Press, N.Y., pp. 77-127.
31. Lathe, R. (1985). J. Mol. Biol. 183, 1-12.
32. Rao, D.R., Beyreuther, K. and Jaenicke, L. (1983). Europ. J. Biochem. 35, 582-592.
33. Turner, N.E., Robinson, S.J. and Haselkorn, R. (1983). Nature 306, 337-342.
34. Cullimore, J.V., Gebhardt, C., Saarelainen, R., Miflin, B.J., Idler, K.B. and Barker, R.F. (1984). J. Mol. App. Gen. 2, 621-635.

35. Donn, G., Tischer, E., Smith, J.A. and Goodman, H.M. (1984). J. Mol. App. Gen. 2, 589-599.
36. Sternberg, M.J.E. and Taylor, W.R. (1984). Febs Letts. 175, 387-392.
37. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982). EMBO J. 1, 945-951.
38. Powers, S.G. and Riordan, J.F. (1975). Proc. Natl. Acad. Sci., USA 72, 2616-2620.
39. Nyunoya, H. and Lusty, C.J. (1983). Proc. Natl. Acad. Sci., USA 80, 4629-4633.

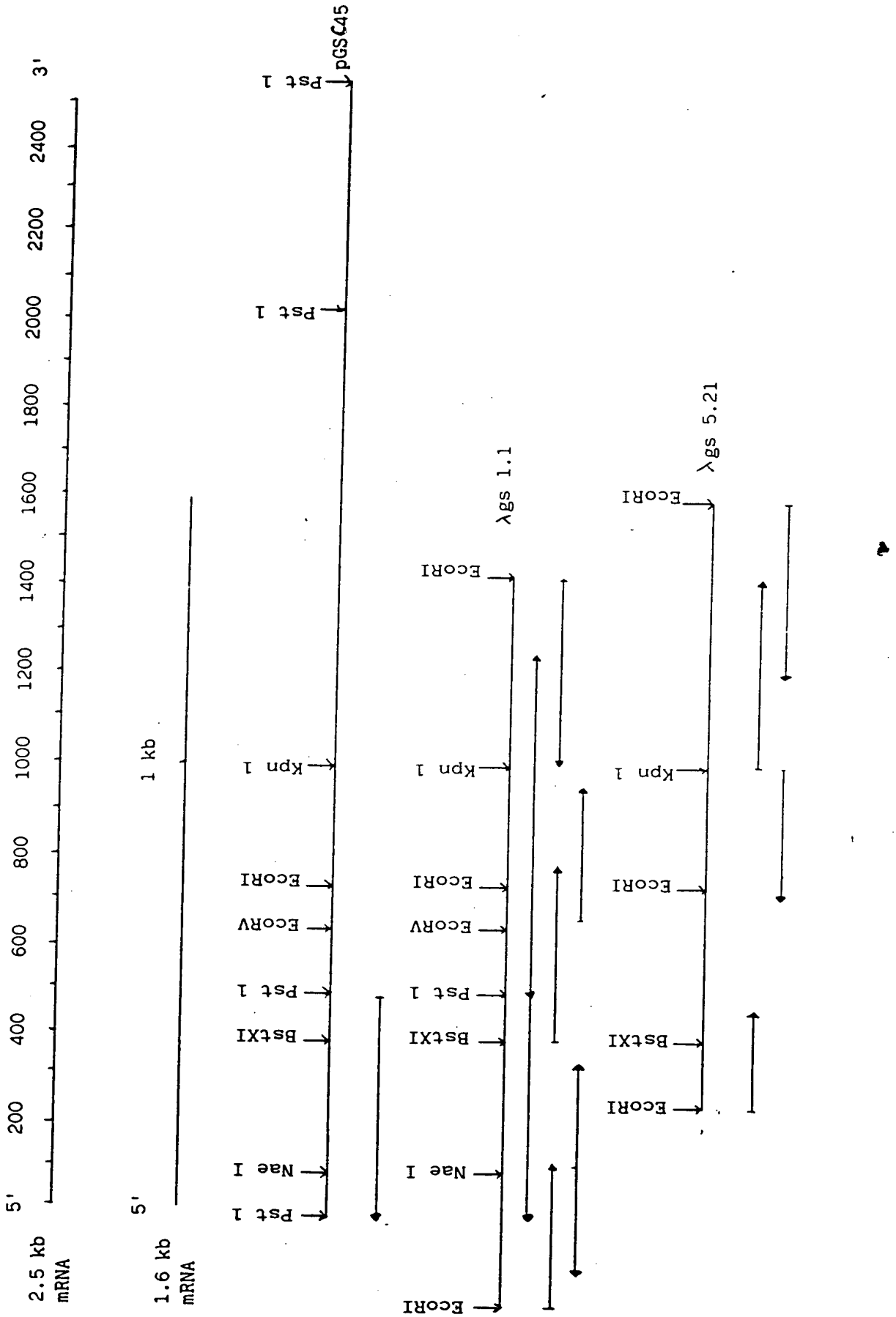


Fig. 1

Fig. 2

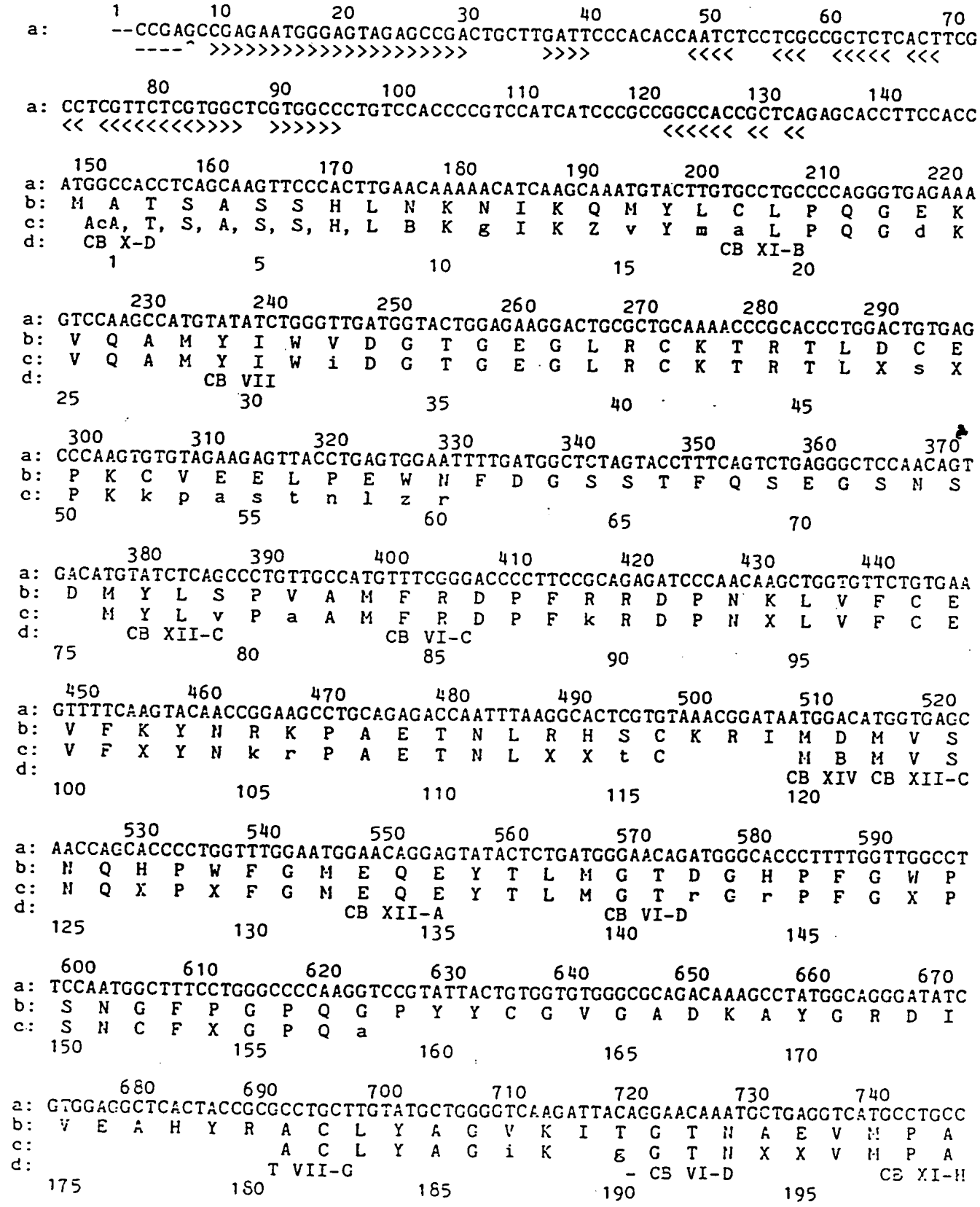

1 10 20 30 40 50 60 70
a: --CCGAGCCGAGAATGGGAGTAGAGCCGACTGCTTGATTCCCACACCAATCTCCTCGCCGCTCTCACTTCG
80 90 100 110 120 130 140
a: CCTCGTTCTCGTGGCTCGTGGCCCTGTCCACCCCGTCCATCATCCCGCCGGCCACCGCTCAGAGCACCTTCCACC
150 160 170 180 190 200 210 220
a: ATGGCCACCTCAGCAAGTTCCCACTTGAACAAAAACATCAAGCAAATGTACTTGTGCCTGCCCCAGGGTGAGAAA
b: M A T S A S S H L N K N I K Q M Y L C L P Q G E K
c: AcA, T, S, A, S, S, H, L B K g I K Z v Y m a L P Q G d K
d: CB X-D CB XI-B
1 5 10 15 20
230 240 250 260 270 280 290
a: GTCCAAGCCATGTATATCTGGGTTGATGGTACTGGAGAAGGACTGCGCTGCAAAACCCGCACCCTGGACTGTGAG
b: V Q A M Y I W V D G T G E G L R C K T R T L D C E
c: V Q A M Y I W i D G T G E G L R C K T R T L X s X
d: CB VII
25 30 35 40 45
300 310 320 330 340 350 360 370
a: CCCAAGTGTGTAGAAGAGTTACCTGAGTGGAATTTTGATGGCTCTAGTACCTTTCAGTCTGAGGGCTCCAACAGT
b: P K C V E E L P E W N F D G S S T F Q S E G S N S
c: P K k p a s t n l z r
50 55 60 65 70
380 390 400 410 420 430 440
a: GACATGTATCTCAGCCCTGTTGCCATGTTTCGGGACCCCTTCCGCAGAGATCCCAACAAGCTGGTGTTCTGTGAA
b: D M Y L S P V A M F R D P F R R D P N K L V F C E
c: M Y L v P a A M F R D P F k R D P N X L V F C E
d: CB XII-C CB VI-C
75 80 85 90 95
450 460 470 480 490 500 510 520
a: GTTTTCAAGTACAACCGGAAGCCTGCAGAGACCAATTTAAGGCACTCGTGTAAACGGATAATGGACATGGTGAGC
b: V F K Y N R K P A E T N L R H S C K R I M D M V S
c: V F X Y N k r P A E T N L X X t C M B M V S
d: CB XIV CB XII-C
100 105 110 115 120
530 540 550 560 570 580 590
a: AACCAGCACCCCTGGTTTGGAATGGAACAGGAGTATACTCTGATGGGAACAGATGGGCACCCTTTTGGTTGGCCT
b: N Q H P W F G M E Q E Y T L M G T D G H P F G W P
c: N Q X P X F G M E Q E Y T L M G T r G r P F G X P
d: CB XII-A CB VI-D
125 130 135 140 145
600 610 620 630 640 650 660 670
a: TCCAATGGCTTTCCTGGGCCCCAAGGTCCGTATTACTGTGGTGTGGGCGCAGACAAAGCCTATGGCAGGGATATC
b: S N G F P G P Q G P Y Y C G V G A D K A Y G R D I
c: S N C F X G P Q a
150 155 160 165 170
680 690 700 710 720 730 740
a: GTGGAGGCTCACTACCGCGCCTGCTTGTATGCTGGGGTCAAGATTACAGGAACAAATGCTGAGGTCATGCCTGCC
b: V E A H Y R A C L Y A G V K I T G T N A E V M P A
c: A C L Y A G i K g G T N X X V M P A
d: T VII-G - CB VI-D CB XI-H
175 180 185 190 195

Fig. 2 continued

```
    750        760        770        780        790        800        810        820
a: CAGTGGGAATTCCAAATAGGACCCTGTGAAGGAATCCGCATGGGAGATCATCTCTGGGTGGCCCGTTTCATCTTG
b:  Q  W  E  F  Q  I  G  P  C  E  G  I  R  M  G  D  H  L  W  V  A  R  F  I  L
c:  Q  W  E  F  Q  I  G  P  C  E  G  I  d  M
   200        205        210        215        220

        830        840        850        860        870        880        890
a: CATCGAGTATGTGAAGACTTTGGGGTAATAGCAACCTTTGACCCCAAGCCCATTCCTGGGAACTGGAATGGTGCA
b:  H  R  V  C  E  D  F  G  V  I  A  T  F  D  P  K  P  I  P  G  N  W  N  G  A
   225        230        235        240        245

    900        910        920        930        940        950        960        970
a: GGCTGCCATACCAACTTTAGCACCAAGGCCATGCGGGAGGAGAATGGTCTGAAGCACATCGAGGAGGCCATCGAG
b:  G  C  H  T  N  F  S  T  K  A  M  R  E  E  N  G  L  K  H  I  E  E  A  I  E
c:                               M  X  E  E  N  G  L  K  y  I  E  E  A  I  E
d:                                  CB III-C
   250        255        260        265        270

        980        990       1000       1010       1020       1030       1040
a: AAACTAAGCAAGCGGCACCGGTACCACATTCGAGCCTACGATCCCAAGGGGGGCCTGGACAATGCCCGTGGTCTG
b:  K  L  S  K  R  H  R  Y  H  I  R  A  Y  D  P  K  G  G  L  D  N  A  R  G  L
c:  X  L  St Ks n  i  n  y  q        A  Y  B  P  K
d:    T IX-B-2                        T IX-B-1
   275        280        285        290        295

   1050       1060       1070       1080       1090       1100       1110       1120
a: ACTGGGTTCCACGAAACGTCCAACATCAACGACTTTTCTGCTGGTGTCGCCAATCGCAGTGCCAGCATCCGCATT
b:  T  G  F  H  E  T  S  N  I  N  D  F  S  A  G  V  A  N  R  S  A  S  I  R  I
c:                T  S  N  I  N  y  q                          g  A  S  I  R  I
d:                   (-CB III-C)                                T IX-C-1
   300        305        310        315        320

        1130       1140       1150       1160       1170       1180       1190
a: CCCCGGACTGTCGGCCAGGAGAAGAAAGGTTACTTTGAAGACCGCCGCCCCTCTGCCAATTGTGACCCCTTTGCA
b:  P  R  T  V  G  Q  E  K  K  G  Y  F  E  D  R  R  P  S  A  N  C  D  P  F  A
c:  P  R
d: T IX-E
   325        330        335        340        345

   1200       1210       1220       1230       1240       1250       1260       1270
a: GTGACAGAAGCCATCGTCCGCACATGCCTTCTCAATGAGACTGGCGACGAGCCCTTCCAATACAAAAACTAATTA
b:  V  T  E  A  I  V  R  T  C  L  L  N  E  T  G  D  E  P  F  Q  Y  K  N ***
c:                      T  C  L  L  N  Z  T  G  B  Z  P  F  Q  Y  K
d:                         T VI-K
   350        355        360        365        370    372

        1280       1290       1300       1310       1320       1330       1340
a: GACTTTGAGTGATCTTGAGCCTTTCCTAGTTCATCCCACCCCGCCCCAGCTGTCTCATTGTAACTCAAAGGATGG

   1350       1360       1370       1380       1390       1400       1410       1420
a: AATATCAAGGTCTTTTTATTCCTCGTGCCCAGTTAATCTTGCTTTTATTGGTCAGAATAGAGGAGTCAAGTTCTT
e: AATATCAAGGTCTTTTTATTCCTCGTGCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 3

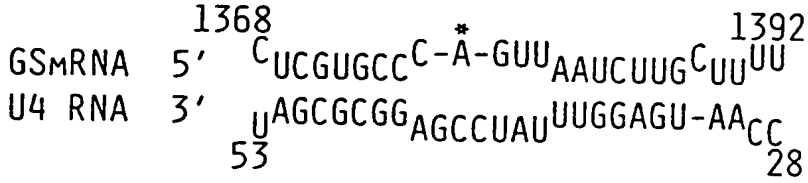


Fig. 4

```
   1 --CCGAGCCG AGAATGGGAG TAGAGCCGAC TGCTTGATTC CCACACCAAT CTCCTCGCCG
  61 CTCTCACTTC GCCTCGTTCT CGTGGCTCGT GGCCCTGTCC ACCCCGTCCA TCATCCCGCC
 121 GGCCACCGCT CAGAGCACCT TCCACCATGG CCACCTCAGC AAGTTCCCAC TTGAACAAAA
 181 ACATCAAGCA AATGTACTTG TGCCTGCCCC AGGGTGAGAA AGTCCAAGCC ATGTATATCT
 241 GGGTTGATGG TACTGGAGAA GGACTGCGCT GCAAAACCCG CACCCTGGAC TGTGAGCCCA
 301 AGTGTGTAGA AGAGTTACCT GAGTGGAATT TTGATGGCTC TAGTACCTTT CAGTCTGAGG
 361 GCTCCAACAG TGACATGTAT CTCAGCCCTG TTGCCATGTT TCGGGACCCC TTCCGCAGAG
 421 ATCCCAACAA GCTGGTGTTC TGTGAAGTTT TCAAGTACAA CCGGAAGCCT GCAGAGACCA
 481 ATTTAAGGCA CTCGTGTAAA CGGATAATGG ACATGGTGAG CAACCAGCAC CCCTGGTTTG
 541 GAATGGAACA GGAGTATACT CTGATGGGAA CAGATGGGCA CCCTTTTGGT TGGCCTTCCA
 601 ATGGCTTTCC TGGGCCCCAA GGTCCGTATT ACTGTGGTGT GGGCGCAGAC AAAGCCTATG
 661 GCAGGGATAT CGTGGAGGCT CACTACCGCG CCTGCTTGTA TGCTGGGGTC AAGATTACAG
 721 GAACAAATGC TGAGGTCATG CCTGCCCAGT GGGAATTCCA AATAGGACCC TGTGAAGGAA
 781 TCCGCATGGG AGATCATCTC TGGGTGGCCC GTTTCATCTT GCATCGAGTA TGTGAAGACT
 841 TTGGGGTAAT AGCAACCTTT GACCCCAAGC CCATTCCTGG GAACTGGAAT GGTGCAGGCT
 901 GCCATACCAA CTTTAGCACC AAGGCCATGC GGGAGGAGAA TGGTCTGAAG CACATCGAGG
 961 AGGCCATCGA GAAACTAAGC AAGCGGCACC GGTACCACAT TCGAGCCTAC GATCCCAAGG
1021 GGGGCCTGGA CAATGCCCGT GGTCTGACTG GGTTCCACGA AACGTCCAAC ATCAACGACT
1081 TTTCTGCTGG TGTCGCCAAT CGCAGTGCCA GCATCCGCAT TCCCCGGACT GTCGGCCAGG
1141 AGAAGAAAGG TTACTTTGAA GACCGCCGCC CCTCTGCCAA TTGTGACCCC TTTGCAGTGA
1201 CAGAAGCCAT CGTCCGCACA TGCCTTCTCA ATGAGACTGG CGACGAGCCC TTCCAATACA
1261 AAAACTAATT AGACTTTGAG TGATCTTGAG CCTTTCCTAG TTCATCCCAC CCCGCCCCAG
1321 CTGTCTCATT GTAACTCAAA GGATGGAATA TCAAGGTCTT TTTATTCCTC GTGCCCAGTT
1381 AATCTCGCCT TTTATTGGTC AGATAGAGGA TCA
```